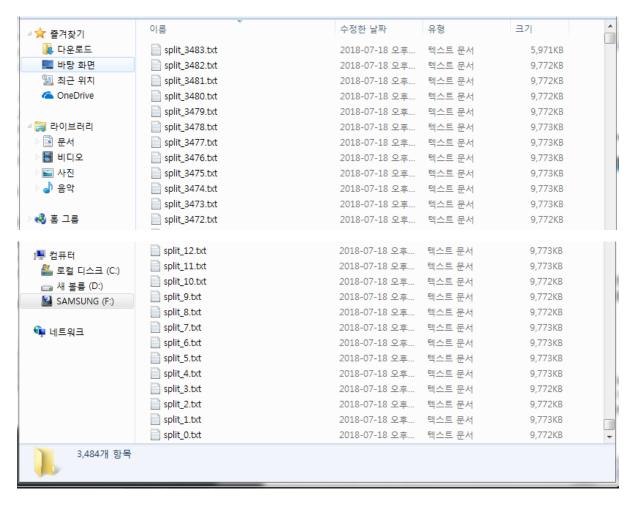
# 방화벽 분석

## BOB 7TH 포렌식 김성수

## 1. Parsing & Inserting

- 1.1 splitData.py
- 1.1.1. Purpose

Firewall.log (120GB) 를 10MB 크기의 12000 개의 파일로 분리한다. 이와 동시에 날짜별 로그를 '\n' 을 통해 구분한다. 다음의 사진은 3484 개, 약 30GB 까지 분리한 상태이다.

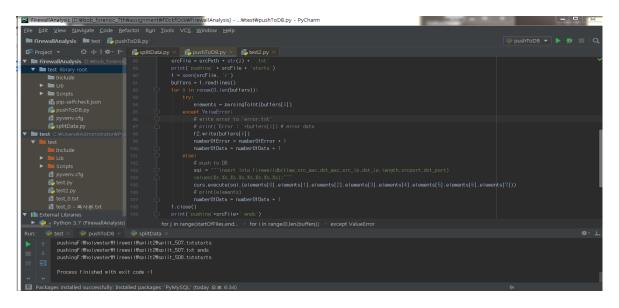


## 1.1.2. Algorithm

- 1) 10KB 단위로 끊어서 문자열을 읽는다.
- 2) '2018-' 을 찾는다.
- 3) 첫번째로, 매칭된 2018- 부터 문자열 끝까지를 새로운 문자열로 정의
- 4) 문자열 내에서, 첫번째 dst\_port 를 찾고, dst\_port="\*" 뒤에 오는 ' ', spacebar 를 '\n'으로 치환
- 5) 반복해서 문자열 내의 dst\_port 를 찾고, 그 뒤의 '', spacebar 를 'n'으로 치환
- 6) 위의 과정을 1000 번 수행해서 10MB 단위로 파일을 생성한다.(ex. split 3121.txt)

### 1.2 pushToDB.py

- 1) '\n' 으로 구분된 로그를, f.readline() 함수를 통해 불러온 뒤, 필요한 항을 뽑는다.
- 2) 각각의 항을, 포맷에 맞게 변경한다.
- 2-1) 첫번째 파일을 읽었을 때, mac address 에 '.', ',' 가 포함되어 있는 것을 확인하고 이를 제거
- 2-2) Mac address 의 경우 hex to decimal 로 치환한다.
- 3) DB 에 저장, data type 은 bigint 이다.(int[11] 일 경우, 시간정보가 들어가지 않음)
- 다음 그림은, 500 개의 파일(5 GB)을 DB 에 저장했을 때의 스냅샷이다.



#### 1.3 error.txt

치환하는데 있어서 에러가 나는 ValueError 의 경우, exception 처리를 통해 error.txt 에 쓴다.

## 2. Infer from error & Query to DB

#### 2.1 Find clue

약 30GB 까지 (10MB 파일 3000 개) 수행하였을 때, error.txt 에 다음의 3 개의 로그가 생긴다.



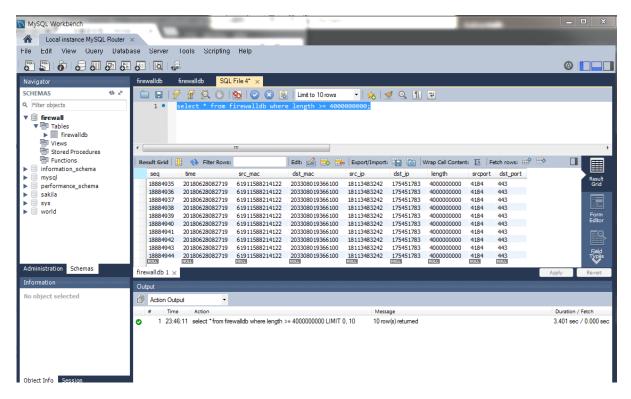
20180628082719/61911588214122/203308019366100/18113483242/175451783/4000000000/4184/443 에서, mac address 를 hex to decimal 하면 다음과 같다.

20180628082719/384EEA39156A/B8E8564638D4/18113483242/175451783/4000000000/4184/443

공격자는 특정서버로 대량의 데이터를 유출했으므로, 위의 로그 정보에서 length 가, 4000000000 로 굉장히 큰 것으로 보아, 이 로그가 공격자에 관한 로그일 것을 의심이 된다.

#### 2.2 Search from DB

이전 단계에서 의심이 되는 로그 파일에 관하여, 더 많은 정보를 얻기 위해, length >= 40000000000으로 DB에 질의를 하였고, 다음 그림과 같은 결과를 얻었다.



### 3. Conclusion

공격자는 src\_mac=38:4E:EA:39:15:6A, src\_ip=18113483242 서버로부터, dst\_mac= B8:E8:56:46:38:D4, dst\_ip=175.45.17.83(175.45.178.3) 공격대상 PC 로 대용량의 데이터를 유출하였음을 알 수 있다.

## 4. DB Dump

